10К

УДК 004.932.751

Ю.В. Крак, Д.В. Шкільнюк
Інститут кібернетики ім. В.М. Глушкова НАН України, м. Київ, Україна
krak@unicyb.kiev.ua, dimonshk@gmail.com

Технологія розпізнавання елементів дактильно-жестової мови

Об’єктом дослідження є дактильно-жестова мова, яка використовується для спілкування людьми з пониженим слухом. Метою дослідження є розробка та реалізація алгоритмів розпізнавання дактильно-жестової мови. У роботі розглядається та використовується метод анімації персонажів, за допомогою якого виокремлюється ряд ознак жестів. Розглядаються алгоритми візуальної обробки зображень. Запропоновано класифікацію жестів дактильної мови.

Вступ

Згідно з останніми даними світових досліджень, близько 2% людей мають ті або інші порушення слуху. На жаль, з кожним роком ця цифра зростає. У свою чергу, порушення слуху породжує проблему розвитку мови. Більшість глухих є інвалідами з дитинства. Навчаючись у спецінтернатах, відвідуючи свої клуби та укладаючи шлюби у власному середовищі, вони утворюють своєрідну субкультуру. Між собою такі особи спілкуються спеціальною жестовою мовою. У свою чергу, їм досить важко контактувати з оточуючим світом. В Україні на державному рівні [1] регулюються питання соціального захисту прав інвалідів, зокрема, «…інвалідам по слуху забезпечується доступ до засобів масової інформації шляхом титрування та здійснення сурдоперекладу на дактильно-жестову мову глухих інформаційних і тематичних телепрограм, кінофільмів, відеофільмів…». Для забезпечення даного закону виникає необхідність у створенні нових засобів спілкування на основі сучасних інформаційних технологій, зокрема, створення комп’ютерних систем моделювання та розпізнавання дактильно-жестової мови.

Отже, об’єктом дослідження є дактильно-жестова мова, якою спілкуються люди з пониженим слухом, а **метою даної роботи** є створення комп’ютерної системи для розпізнавання дактильно-жестової мови.

Постановка задачі дослідження. Дактильно-жестова мова зображує букви алфавіту, цифри, слова і словосполучення. Люди з вадами слуху використовують дактильну мову для відображення власних назв, імен, іншомовних, наукових термінів, речовин і т.д.

Також існує поняття калькулятивної мови жестів, яка використовується при офіційному та діловому спілкуванні і містить в собі як знаки розмовної мови, так і знаки дактильної абетки. Зазначимо, що дактильна абетка слугує для відображення слова по буквах. У свою чергу, калькулятивна мова підпорядковується граматичним правилам державної мови [2].

Деякі автори для ідентифікації елементів зображення використовують нейронну мережу, побудовану за схемою багатошарового персептрона, навчання якої виконано за модифікованим методом зворотного поширення похибки [3], [4]. Експериментально підібрано параметри цієї мережі – один прихований шар з п’ятьма нейронами та непарною активаційною функцією. Для уточнення синаптичних зв’язків мережі алгоритм

навчання використовує відмінність навчального прикладу від побудованого еталона. Задачу класифікації вирішено у два етапи: на першому з них на навчальних прикладах навчено мережу, а на другому – таку мережу використано для класифікації нових прикладів. У роботах [5], [6] пропонується застосовувати приховану модель Маркова, в роботі [7] розробляється система розпізнавання жестів, в якій використовується сегментація зображення, відстеження обличчя і рук. Для вдалого відстеження обличчя і рук використовуються алгоритми виявлення шкіри, особливостей руху, позицій рук і обличчя. Після визначення руки проводиться аналіз форми і порівняння з деякими еталонними зображеннями. В роботі [8] пропонується встановлювати камери перпендикулярно до столу, так щоб в поле зору камери потрапляли лише стіл і руки, і, отже, відпадає потреба реалізовувати дії для розрізнення обличчя людини і її рук.

У техніці часто використовують ехолокацію для визначення положення об'єкта. Ехолокація – спосіб, за допомогою якого положення об'єкта визначається за часом затримки відбитої хвилі. Робота ехолокації може базуватися на відбитті сигналів різної частоти – радіохвилі, ультразвук і звук. Деякі ехолокаційні системи направляють сигнал у визначену точку простору і за затримкою відповіді визначається її віддаленість при відомій швидкості переміщення даного сигналу в даному середовищі і властивості об'єкта, до якого визначається відстань, відбивати даний вид сигналу. Але дослідження положення об'єктів за допомогою даного методу вимагає часових затрат. Також використовуються технічні системи з одночасним використанням сигналів різної частоти, які дозволяють істотно прискорити процес ехолокації.

Ехолокаційну техніку можна застосовувати при отриманні інформації про жести. Зокрема, розглянемо випадок, коли людина стоїть фронтально до ехолокаційної системи. Проаналізувавши особливості виконання деякого жесту, можна зробити наступні висновки: що рука, якою виконується жест, завжди буде ближче до системи ніж інша частина тіла. Тому, необхідно досліджувати область, яка знаходиться найближче до ехолокаційної системи. У свою чергу, в даній області будуть зони, які знаходяться ближче або дальше від ехолокаційної системи. На основі отриманих даних будується об'ємне зображення руки, якою виконується жест, і досліджуються особливості виконання жесту.

Метод лазерного сканування дозволяє створювати цифрову модель оточуючого простору, представляючи його у вигляді набору точок з просторовими координатами [9]. Схематично лазерний сканер можна розділити на декілька основних блоків: 1) головка вимірювання. Як правило, в ній розташований лазерний випромінювач і приймач; 2) обертаюча призма. Забезпечує розподілення пучка у вертикальній площині; 3) привід горизонтального круга. Забезпечує обертання головки вимірювання в горизонтальній площині; 4) комп'ютерна система. Призначена для керування зйомкою і запису даних на носій.

Отримання координат точок об'єкта лазерним сканером базується на вимірюванні полярних кутів і відстаней до об'єкта. Напрям вимірювання визначається оптико-механічною системою сканера. На шляху до об'єкта імпульси лазерного випромінювання проходять через систему дзеркал, які виконують покрокове відхилення променя. Найбільш розповсюдженою є конструкція, яка складається з двох рухомих дзеркал. Одне з них відповідає за вертикальне переміщення променя, друге – за горизонтальне переміщення дзеркал. У кінцевому результаті дзеркала забезпечують точність напрямлення променя лазера на об'єкт. Дзеркала відхиляються на задану величину і таким чином дозволяють визначити полярний кут, який необхідний для находження координат. Оскільки рука, якою виконується жест, завжди буде ближче

до системи ніж інша частина тіла, то при використанні лазерного сканера також досліджуються області, які находяться найближче до системи. Як і при використанні ехолокаційної системи, на основі отриманих даних будується об'ємне зображення руки, якою виконується жест, і досліджуються особливості виконання жесту.

Для чіткого виявлення ознак жестів використовують технологію захвату рухів (MoCap) [10]. Даний метод застосовується для зйомок мультфільмів, для створення спеціальних ефектів в художніх фільмах, а також широко використовується в ігровій індустрії. Система MoCap використовує спеціальне обладнання (камери), зокрема, костюми з датчиками або світодіодами.

На сьогоднішній день існує велика кількість систем захвату руху, різниця між якими полягає в принципі передачі руху. Так, дані системи поділяються на дві групи:

1. Оптичні пасивні. Основною характерною ознакою таких систем є прикріплені до костюма датчики-маркери, які називаються пасивними, тому що відбивають світло, яке попадає на них, а самі вони, у свою чергу, не світяться. У таких системах світло (інфрачервоне) на маркери посилається з встановлених на камерах високочастотних стробоскопів і, відбившись від маркерів, попадає знову в камеру, визначаючи цим позицію маркера. Недоліком таких систем є те, що при швидкому русі або близькому розміщенні маркери плутаються або зміщуються, тобто система може сплутати порядок маркерів.

2. Оптичні активні. Дані системи характеризуються тим, що замість світловідбиваючих маркерів використовуються світлодіоди з інтегрованими процесорами і радіосинхронізацією. Кожному світлодіоду призначається своя адреса, що дозволяє їх не плутати, а також розпізнавати їх після того, як вони були перекриті і знову з'явилися в полі зору камер. В усьому іншому принцип роботи активних систем подібний до пасивних систем.

Так, людина одягає спеціальний костюм із світодіодами та проводить певні рухи або приймає різноманітні стани. За допомогою програмного забезпечення дані, отримані з камер і датчиків, відтворюються в анімаційному вигляді.

Для отримання даних про рух зап'ястя, пальців використовується рукавичка з світодіодами (рис. 1). Зокрема, людина одягає рукавичку і демонструє окремі рухи. У свою чергу, камери і датчики збирають дані, програмне забезпечення їх обробляє і відтворює рухи в анімаційному вигляді (рис. 2).



Рисунок 1 – Рукавичка з світлодіодами

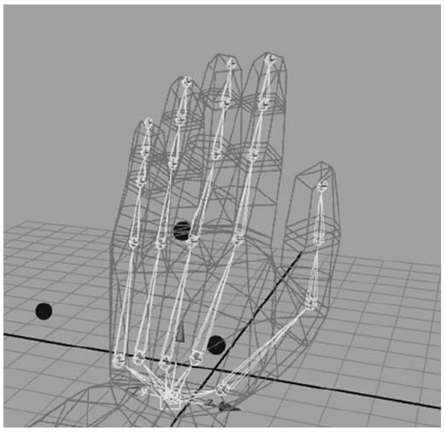


Рисунок 2 – Анімаційне зображення рухів руки

У даній статті запропоновано ряд алгоритмів для ідентифікації елементів жестів, в яких використовується попіксельний аналіз зображення і елементів жесту. Також досліджені методи і алгоритми збору ознак, за якими буде відбуватися розпізнавання і класифікація жестів дактильної мови.

Використання веб-камери і деяких алгоритмів обробки зображення

Для вивчення природи жестів використовується звичайна веб-камера і рукавички, колір яких різко відрізняється від кольорів предметів, що попадають у поле зору веб-камери. Перед початком роботи даного методу необхідно детально вивчити природу жестів, які відображають букву. Розпізнавання жестів можна визначити як відношення вхідних даних до певного класу образів за допомогою виділення істотних ознак або властивостей, котрі виділяють вхідні дані від загальної маси неістотних деталей. Під класом образів розуміється певна категорія, яка визначається рядом ознак, спільних для всіх її елементів [11]. Отже, жести, які відображають букви, можна поділити на дві групи: рухомі і нерухомі. Жест на відеозображенні розглядається як послідовність кадрів, на яких відбувається зміна форми кисті руки [3]. До рухомих жестів відносяться жести, які відображають наступні букви: Д, Є, З, Ї, Й, К, Ф, Х, Ц, Щ, Ю, Я, Ь. До нерухомих: А, Б, В, Г, Е, И, Ж, І, Л, М, Н, О, П, Р, С, Т, У, Ч, Ш.

Розглянемо випадок, коли в поле зору камери попадає жест, який відображає нерухому букву. Побудуємо квадрат, розміри якого дозволяють охопити руку з розкритими пальцями. Бажано, щоб сторона квадрата дорівнювала 100 пікселям. Даний квадрат необхідно розбити на рівні клітини, сторони яких будуть рівні 25 пікселям (рис. 3).

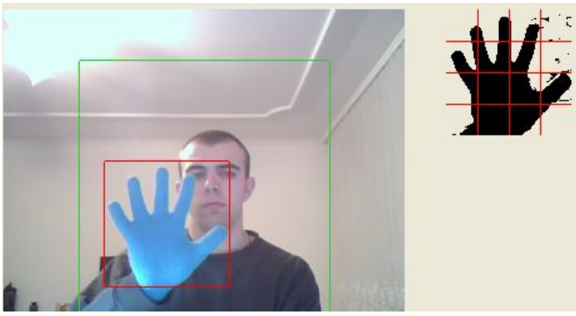


Рисунок 3 – Побудова необхідного квадрата

Після проведених дій необхідно провести показ кожного жесту, який відображає нерухому букву. Для кожного жесту проводиться аналіз – в які клітини попадає рука при відтворенні жесту. Дані клітини замальовуються цілковито (рис. 4).

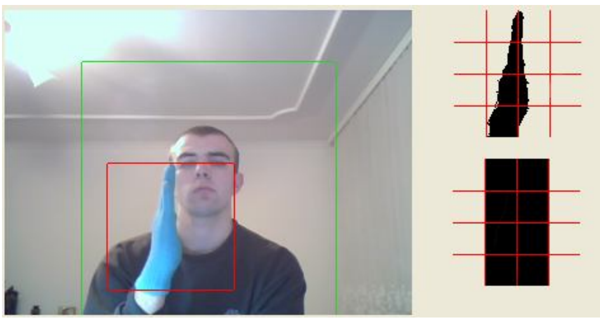


Рисунок 4 – Аналіз розміщення руки в квадраті

У кожній клітині обраховується відношення кількості чорних точок до загальної кількості точок. Тобто будуємо значення еталонів для кожного жесту [11]. В даному випадку еталон буде містити шістнадцять значень. Сукупність еталонів жестів утворюють базу знань, яка в наступних кроках буде використовуватись для розпізнавання.

Провівши послідовність дій для кожного жесту, можна виділити наступні класи образів :

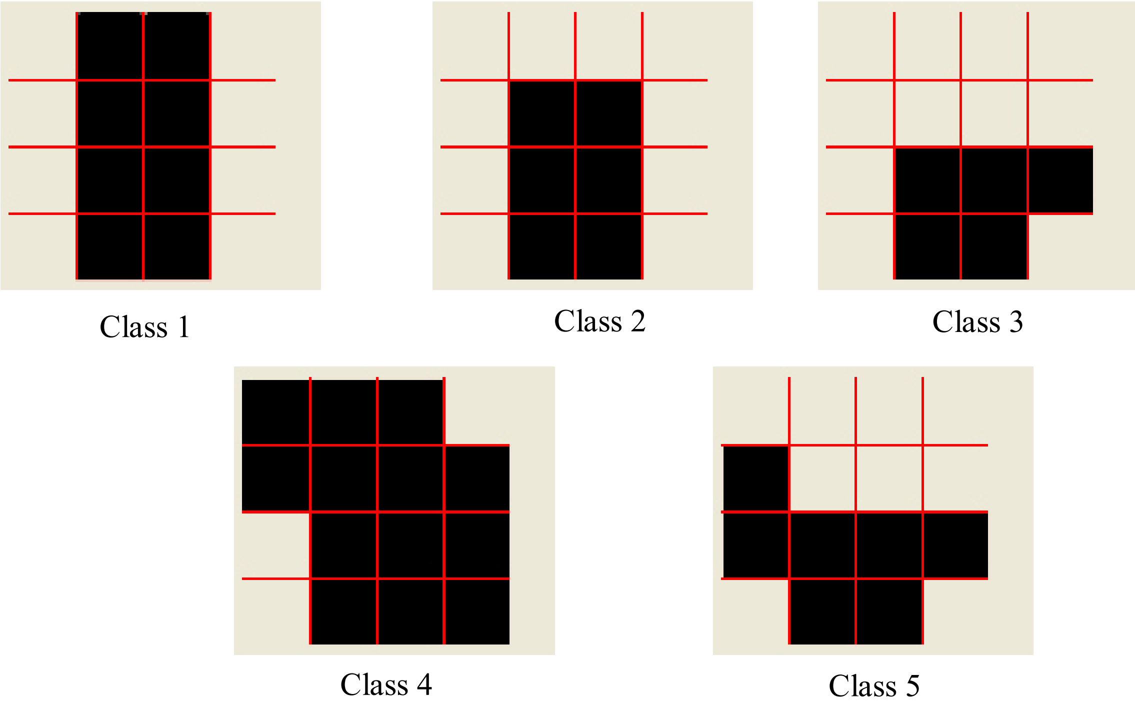


Рисунок 5 – Класи образів

Жести, які відтворюють нерухомі букви, розділені між класами наступним чином:
Class 1 – Б, В;
Class 2 –А, Г, І, Л, М, Ч, П, Ш;
Class 3 – Ж, Ч, Е;
Class 4 – О, Р, С, Т;
Class 5 – У.

Під час розпізнавання жестів, що відображають нерухомі букви, в першу чергу визначається, до якого класу відноситься даний жест. Наступний крок – це побудова еталона вхідного жесту і порівняння з еталонами жестів (з бази знань) даного класу. Далі необхідно обчислити суму різниць між значеннями клітин еталонів і значенням клітин жесту, що розпізнається.

$$\sum_{i,j}^{n} \left| a_{ij} - M_{ij}^{k} \right| = A^{k},$$

де a_{ij} – значення кожної клітини жесту, який розпізнається, M_{ij}^{k} – значення клітини кожного еталона. З множини A^k вибирається найменше. Найменше A^k визначить еталон, до якого відноситься жест, що розпізнається.

Під час виконання вищенаведеного прикладу виникають деякі проблеми при визначенні положення окремих пальців. Даний недолік можна виправити за допомогою рукавички з різними кольорами пальців. Використання різних кольорів для

пальців надає можливість точніше визначити положення окремого пальця і приблизно визначити, в якому стані палець – зігнутий чи розігнутий.

Так, кожному пальцю відповідає окремий колір:
– жовтий – великий палець;
– червоний – вказівний палець;
– зелений – середній палець;
– синій – безіменний палець;
– фіолетовий – мізинець.

Розглянемо даний метод для обробки жесту, який відображає букву «Р» (рис. 6), і порівняємо з попереднім методом (рис. 7).

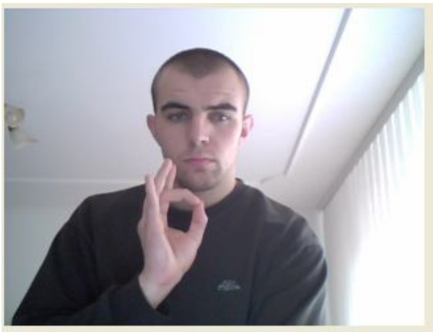


Рисунок 6 – Буква «Р» у звичайному зображенні

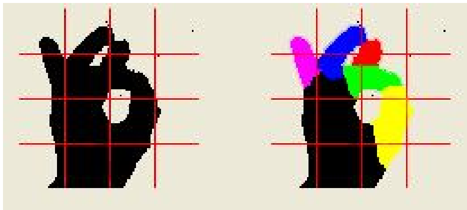


Рисунок 7 – Порівняння методів

Вищенаведені алгоритми вдало працюють при сталій фокусній відстані між камерою і рукою.

У випадку, коли в поле зору камери потрапляє жест з великими амплітудними рухами, необхідно відслідкувати траєкторію руху руки. Для відстеження траєкторії об’єкта необхідно слідкувати за координатами центра мас. Координати центра мас об’єкта обраховуються наступним чином.

Нехай N – кількість пікселів, які відносяться до об’єкта. Всю множину пікселів $p(x, y)$, що відносяться до об’єкта, позначимо Ω [12]. Тоді координати центра мас обчислюються наступним чином:

$$x_c = \frac{1}{N}\sum_{p(x,y)\in\Omega} x,\; y_c = \frac{1}{N}\sum_{p(x,y)\in\Omega} y .$$

Проаналізувавши траєкторії рухомих жестів, букви, які відображаються за допомогою даних жестів, пропонуємо розділити на наступні класи:
– «За годинниковою стрілкою» – Д (жести, траєкторія яких іде за годинниковою стрілкою фронтально до співрозмовника);
– «Навколо своєї осі» – Є, Ї, Й, К, Ф, Ь (жести, траєкторія яких іде навколо осі руки);
– «Зліва направо» – З, Я (жести, траєкторія яких іде зліва направо фронтально до співрозмовника);
– «Справа наліво» – Х, Ю (жести, траєкторія яких іде справа наліво фронтально до співрозмовника);
– «Зверху вниз» – Ч, Ц, Щ (жести, траєкторія яких іде зверху вниз).

Зміни координат центра мас відображаються графічно, що дасть можливість чіткіше визначити чи жест рухомий, чи не рухомий, а також визначити, до якого класу відноситься рухомий жест.

Розглянемо роботу даного методу на прикладі слова «ТРОЯ».

Слово «ТРОЯ» складається з чотирьох букв. Букви «Т», «Р», «О» відображаються нерухомими жестами, а буква «Я» відображається рухомим (рис. 8).

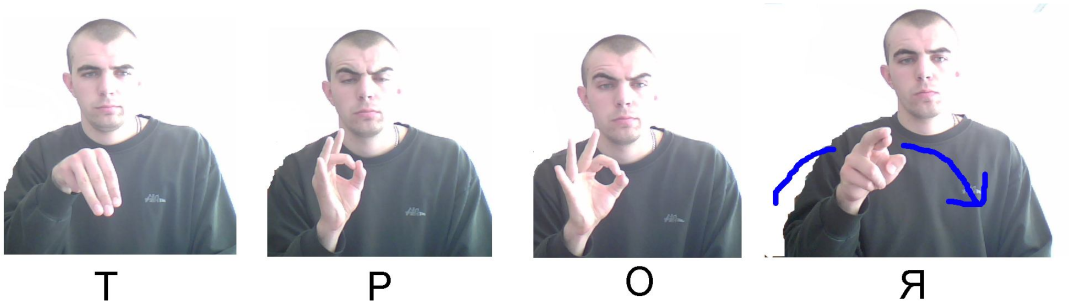


Рисунок 8 – Слово «Троя»

Як було сказано, зміни координат центра мас варто відображати графічно, тому побудуємо два графіка. На вісі абсцис відображається моменти часу, в який відображається жест. На вісі ординат відображаються координати x і y, відповідно на першому і на другому графіках (рис. 9).

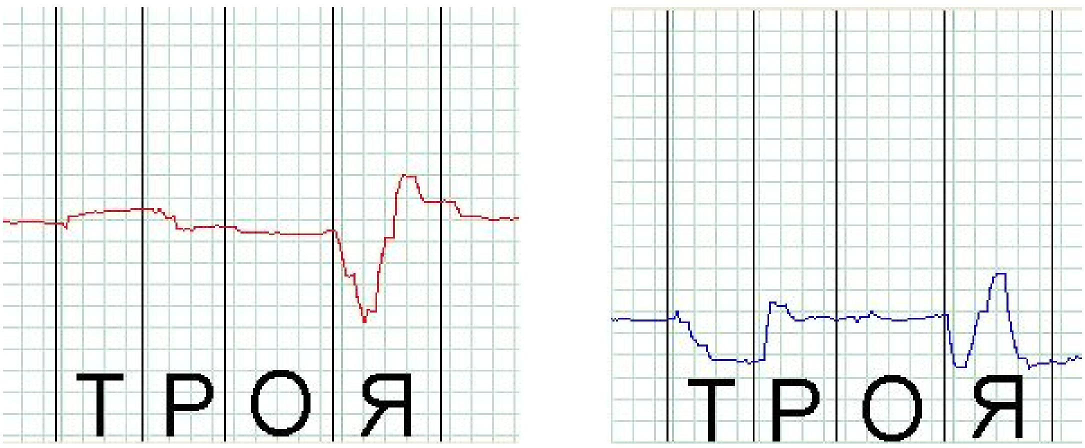


Рисунок 9 – Графічне зображення зміни координат центра мас

Проаналізувавши графіки, можна визначити, де нерухомі жести і жести з великими амплітудними рухами.

Розглянемо приклад, коли в слові присутні декілька букв з великими амплітудними рухами. Наприклад, слово ЮЛЯ, в якому буква Л відображається нерухомим жестом, а букви Ю, Я відображаються рухомими (рис. 10).

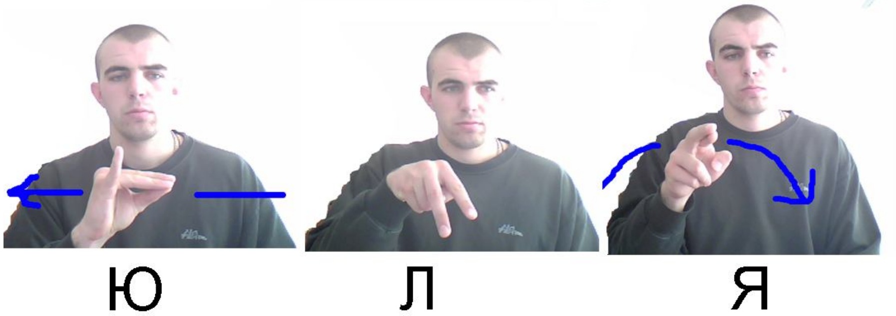


Рисунок 10 – Слово ЮЛЯ

Відслідковуючи рух руки, можна визначити зміну координат центра мас пікселів (рис. 11).

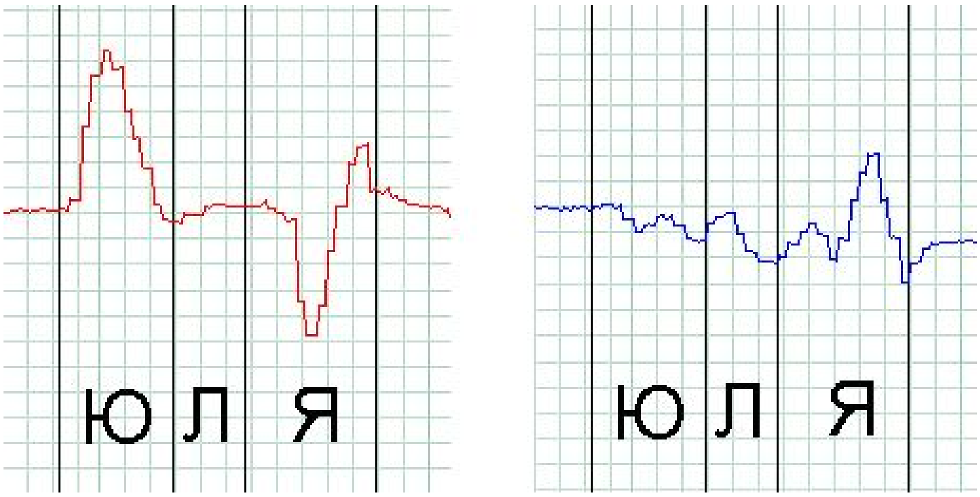


Рисунок 11 – Графічне зображення зміни координат центра мас для слова ЮЛЯ

Таким чином, запропоновану класифікацію жестів можна відобразити за допомогою наступної схеми (рис. 12):

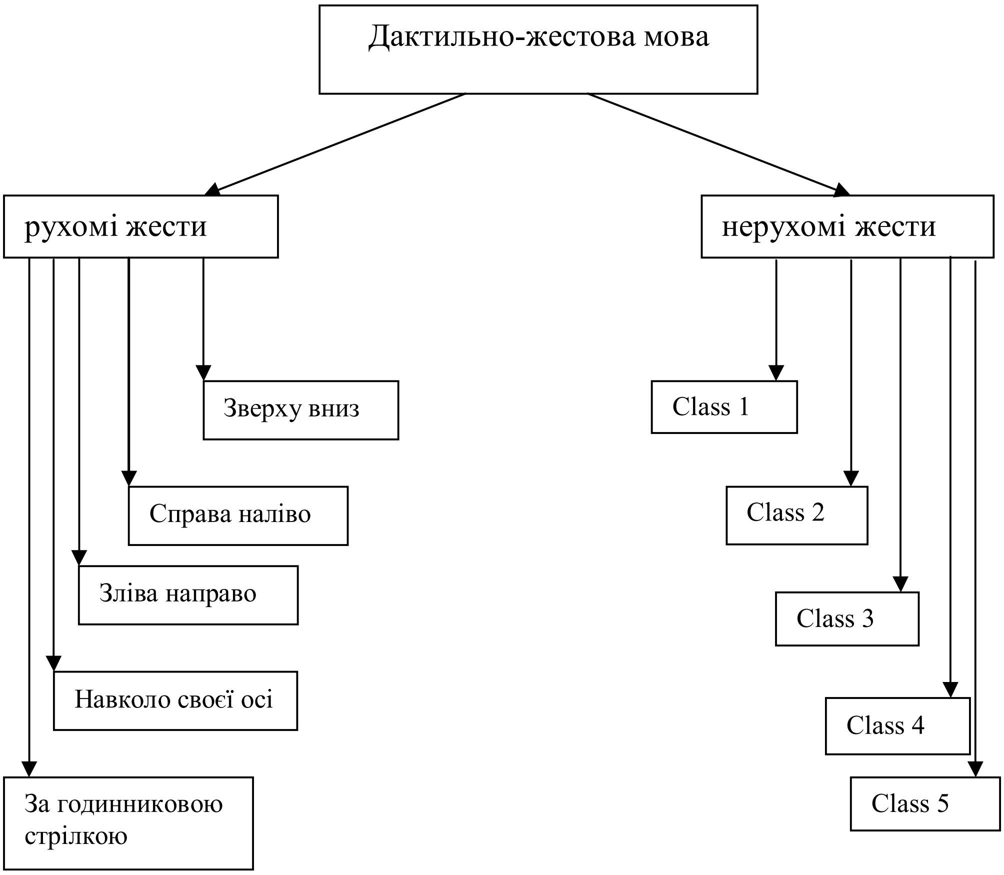


Рисунок 12 – Класифікація дактильних жестів

Висновки

У даній роботі було розглянуто загальну класифікацію жестів дактильно-жестової мови. Описано розпізнавання жестів різних груп, а також методи збору ознак для розпізнавання. Окремо розглянуто принципи роботи системи при розпізнаванні нерухомих жестів і жестів з великими амплітудними рухами. При роботі модуля для розпізнавання жестів важливу роль відіграє графічне відображення руху центра мас, що дасть можливість чітко визначити чи жест нерухомий, чи з великим амплітудним рухом. І якщо жест з великим амплітудним рухом, то графічне зображення дасть можливість визначити, до якого класу належить жест.

На основі проведених випробувань і отриманих результатів розробляються алгоритми розпізнавання дактильно-жестової мови при відсутності рукавичок, або при відсутності допоміжних міток на руках.

Література

1. Закон України «Про основи соціальної захищеності інвалідів в Україні».
2. Крак Ю. Компьютерная система виртуального общения людей с проблемами слуха / Ю. Крак, А. Бармак, А. Ганжа, А. Тернов, Н. Шатковский // 16th International Conference «Knowledge-Dialogue-Solution» KDS 2008. – Varna, Bulgaria, 2008. – С. 161-165.
3. Давидов М.В. Методи та засоби опрацювання зображень реального часу для ідентифікації елементів жестової мови / М.В. Давидов, Ю.В. Нікольський // Штучний інтелект. – 2008. – № 1. – С. 131-138.
4. Давидов М.В. Класифікація елементів відеозображень реального часу з допомогою нейромережі / М.В. Давидов, Ю.В. Нікольський // Вісник Національного університету «Львівська політехніка»: Інформаційні системи та мережі. – 2005. – № 549. – С. 82-92.
5. Heung-Il Suk Robust Modelling and Recognition of Hand Gestures with Dynamic Bayesian Network / Heung-Il Suk, Bong-Kee Sin, Seong-Whan Lee // Pattern Recognition, 2008. ICPR 2008. 19th International Conference, (8-11 Dec. 2008) – P. 1-4.
6. Marcel Sebastien. Hand Gesture Recognition using Input-Output Hidden Markov Models / Sebastien Marcel, Olivier Bernier, Jean-Emmanuel Viallet, Daniel Collobert // Face and Gesture Recognition (FG '00). – P. 456-462.
7. Thomas Coogan. Real time hand gesture recognition including hand segmentation and tracking / Thomas Coogan, George Awad, Junwei Han, Alistair Sutherland // Advances in Visual Computing. – 2006. – P. 495-504.
8. Radu-Daniel Vatavu. Above-the-Table Interactions for Intelligent Sensing Systems / Radu-Daniel Vatavu, Stefan-Gheorghe Pentiuc // 9th International Conference on Development and application systems, Suceava, Romania, (May 22-24, 2008). – P. 285-288.
9. Врачева А.А. Анализ методов и геодезических технологий наземного лазерного сканирования : автореф. магистерской работы [Электронный ресурс] / А.А. Врачева. – Режим доступа: http://masters.donntu.edu.ua/2008/ggeo/vracheva/diss/index.htm.
10. Midori Kitagawa. Workflow and Techiques for Motion Capture / Midori Kitagawa, Brian Windsor. MoCap for Artists. – San Francisco : Focal Press is an imprint of Elsevier, 2008. – 216 p.
11. Ту Дж. Принципы распознавания образов / Дж. Ту, Р. Гонсалес. – М. : Мир, 1978. – 411с.
12. [Электронный ресурс]. Режим доступа: http://www.gonka.ru/nick/glaz/.
13. Форсайт Д. Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс. – М. : Вильямс, 2004. – 926 с.

Ю.В. Крак, Д.В. Шкильнюк

Технология распознавания элементов дактильно-жестового языка

В статье исследуется дактильно-жестовый язык, используемый для общения людьми с дефектами слуха и речи. Целью исследования есть разработка системы распознавания дактильно-жестового языка. Рассматривается и используется метод анимации персонажей, с помощью которого выделяется ряд признаков жестов. Предлагается классификация жестов дактильно-жестового языка.

Yu.V. Krak, D.V. Shkilnyuk

Technology of Recognition of Elements of Finger-Sign Language

In the article finger-sign language for deaf people communication is investigated. The purpose is to develop the system of the finger-sign language recognition. The method of characters animation is considered and used which allows to select some features of gestures. Classification of gestures of finger-sign language is offered.

Стаття надійшла до редакції 09.06.2009.